

Serial No. 10/059,421
Docket No. YOR920010126US2

AMENDMENTS TO THE CLAIMS:

1. (Currently amended) A system for identifying genes, comprising:
a pattern database comprising patterns of amino acids;
an input device for inputting a DNA sequence; and
a processor which locates ~~discovers~~ a pattern from said pattern database in said DNA sequence to identify a putative gene in said DNA sequence.
2. (Original) The system according to claim 1, wherein said processor determines an open reading frame (ORF) in said DNA sequence.
3. (Original) The system according to claim 2, wherein said processor generates an amino acid translation for said ORF.
4. (Original) The system according to claim 3, wherein said processor identifies a match of a pattern from said pattern database in said amino acid translation.
5. (Original) The system according to claim 4, wherein said patterns are derived from a parent database comprising at least one amino acid sequence.
6. (Original) The system according to claim 4, wherein said patterns are derived from a parent database comprising at least one amino acid sequence fragment.
7. (Original) The system according to claim 4, wherein said patterns are derived by using a pattern discovery algorithm.

Serial No. 10/059,421

Docket No. YOR920010126US2

8. (Original) The system according to claim 4, wherein said patterns are derived by using the Teiresias algorithm.
9. (Original) The system according to claim 4, wherein said ORF comprises a portion of said DNA sequence between a start codon and a stop codon.
10. (Original) The system according to claim 4, wherein said ORF is reported as a putative gene when a predetermined number of pattern matches is identified in said amino acid translation.
11. (Original) The system according to claim 4, wherein each pattern is assigned a weight depending upon a relevance of said pattern in determining whether said ORF comprises a putative gene.
12. (Original) The system according to claim 4, wherein said ORF is reported as a putative gene when the sum of weights corresponding to all patterns with matches in said amino acid translation exceeds a predetermined threshold.
13. (Original) The system according to claim 4, wherein said match is identified using a predetermined pattern matching algorithm.
14. (Original) The system according to claim 4, further comprising:
a memory device for storing data and instructions to be executed by said processor.
15. (Original) The system according to claim 4, further comprising:
a display device for displaying an output from said processor.

Serial No. 10/059,421

Docket No. YOR920010126US2

16. (Previously presented) A method of identifying genes, comprising:
providing a pattern database comprising patterns of amino acids;
determining an open reading frame (ORF) in a DNA sequence;
generating an amino acid translation for said ORF; and
identifying a putative gene in said DNA sequence by identifying a match of a pattern
from said pattern database in said amino acid translation.
17. (Original) The method according to claim 16, wherein said pattern database is generated
from a database comprising at least one amino acid sequence.
18. (Original) The method according to claim 16, wherein said pattern database is generated
from a database comprising at least one amino acid sequence fragment.
19. (Original) The method according to claim 16, further comprising:
identifying said ORF as a putative gene when a predetermined number of pattern matches
is identified in said amino acid translation.
20. (Original) The method according to claim 16, further comprising:
assigning a weight to each pattern depending upon a relevance of said pattern in
determining whether said ORF is a putative gene.
21. (Original) The method according to claim 16, further comprising:
displaying said match of said pattern in said amino acid translation.
22. (Original) The method according to claim 16, wherein said pattern database is generated
using the Teiresias algorithm to derive said patterns from a parent database.

Serial No. 10/059,421

Docket No. YOR920010126US2

23. (Previously presented) A programmable storage medium tangibly embodying a program of machine-readable instructions executable by a digital processing apparatus to perform a method for identifying genes, said method comprising:

- providing a pattern database comprising patterns of amino acids;
- determining an open reading frames (ORF) in a given DNA sequence;
- generating an amino acid translation for each ORF; and
- identifying a putative gene in said DNA sequence by identifying a match of a pattern from said pattern database in said amino acid translation.

24. (Currently amended) The system according to claim 3, wherein said processor determines for each pattern in said pattern database whether the pattern is present in said amino acid translation by locating ~~iteratively examines said patterns in said pattern database to locate~~ instances of said patterns in said amino acid translation, until a sum of weights corresponding to all patterns with matches in said amino acid translation exceeds a predetermined threshold, at which point said processor identifies said ORF as a putative gene.

25. (Previously presented) The system according to claim 1, further comprising:

- a parent database comprising a plurality of amino acid sequences, said patterns in said pattern database being derived from said plurality of amino acid sequences by using a pattern discovery algorithm;
- a memory device for storing data and instructions to be executed by said processor; and
- a display device for displaying an output from said processor.

26. (Previously presented) The system according to claim 25, wherein said processor determines an open reading frame (ORF) which comprises a portion of said DNA sequence between a start codon and a stop codon, generates an amino acid translation for said ORF, and identifies a match of a pattern from said pattern database in said amino acid translation by using a

Serial No. 10/059,421
Docket No. YOR920010126US2

predetermined pattern matching algorithm,

wherein each pattern is assigned a weight depending upon a relevance of said pattern in determining whether said ORF comprises a putative gene, and

wherein said ORF is reported as a putative gene when either a predetermined number of pattern matches is identified in said amino acid translation, or a sum of weights corresponding to all patterns with matches in said amino acid translation exceeds a predetermined threshold.

27. (Previously presented) The system according to claim 1, wherein said processor accesses said pattern database to retrieve said patterns from said pattern database.

28. (Previously presented) The system according to claim 1, wherein said processor is electrically coupled to said input device and said pattern database.

29. (Previously presented) A system for identifying genes, comprising:
an input device which inputs a DNA sequence; and
a processor which accesses a pattern database comprising a plurality of patterns of amino acids, and locates instances of said patterns of amino acids in said DNA sequence to identify a putative gene in said DNA sequence.